

A study of the impact of training in a management development program based on 360 feedback

77

Robert T. Rosti Jr

Management & Training Corporation, Ogden, USA, and

Frank Shipper

*Franklin P. Perdue School of Business, Salisbury State University,
Salisbury, USA*

In general, quantitative analysis is rarely done in organizations to ensure that training is effective (Carnevale and Schultz, 1990; Stephen *et al.*, 1988). The same is true also of management training based on multirater or 360 feedback (Church, 1997). Thus, the purpose of this study is to evaluate the effectiveness of management training when done in conjunction with 360 feedback.

When quantitative analyses of management using one-time survey data has been done in the past, it has focused on the frequency of managerial behaviors and not the skill of those behaviors (Bernardin and Beatty, 1984; Schriesheim and Kerr, 1974; Shipper, 1991; Van Velsor and Leslie, 1991; Yukl, 1989, 1994). Recent work has placed an emphasis on managerial skill and not frequency of behavior (e.g. Cox and Cooper, 1988; Luthans *et al.*, 1985; Shipper, 1991; Whitley, 1989). According to one review of 360 feedback instruments and research, skill and not frequency of managerial behaviors may be the important attribute concerning subunit performance (Van Velsor and Leslie, 1991). In addition, Porter and McKibbin (1988) have asserted that training in managerial skills is lacking. Thus, this study will focus on the impact training has on the managerial skills of the participants in a management development program based on 360 feedback.

The most common way that training has been evaluated, when it is done, is through the reaction of the participants to the training (Brown, 1980; Dunn and Thomas, 1985). This assumes that if the participants like the training, it must be effective. Such evaluations provide little substantive information regarding the value of the training. Obviously, a more rigorous assessment of the effectiveness of the training is needed (Brown, 1980; Carnevale and Schultz, 1990; Dunn and Thomas, 1985).

The use of 360 feedback provides a natural method for both enhancing the learning of the participants and improving the evaluation process. Feedback is seen as a critical element in affecting change (Bennis *et al.*, 1969). According to Zemke and Zemke (1995) adults undertake learning experiences when they see a need for a new or different skill or knowledge. When 360 feedback is presented in a modified Johari window, managers can see the skills they need to improve (Shipper and John, 1991). The use of 360 feedback can be the first step, unfreezing current behavior, in Lewin's (1948) model of change which would enhance the learning process. In addition, by conducting 360 feedback before and after the training, the first step in evaluating the effectiveness of the training can occur. In addition, the second round of feedback can be used both as reinforcement of past learning and also an opening for future learning. Moreover, the data from the second round of 360 allows one to examine for positive changes in the subjects' managerial skills as suggested by Kirkpatrick (1976), and for the collection to come from a source other than the manager. By definition, in 360 feedback observations are collected from the employees as well as from the manager, the superior to that manager, and the peers. Multiple studies have shown that the employees' observations of the managers' behaviors have the highest correspondence of all four with performance (Bass, 1990; Clark *et al.*, 1992; Schriesheim and Kerr, 1974; Shore *et al.*, 1992; Wilson *et al.*, 1990). Thus, collecting 360 feedback both before and after the training will both enhance the learning and provide at least part of the data needed to evaluate the training.

An unresolved issue is what behaviors to study. Multiple approaches have been taken to identify the behaviors of effective management. Classical management theorists like Fayol (1949) and Urwick (1952) professed that all management functions involved planning, organizing, co-ordinating, directing, and controlling. These functions require technical, problem-solving, decision-making, and people handling skills. These broad-based categories, however, lack sufficient definitional detail to study managerial proficiency or the effectiveness of training.

The more prominent models of management that have guided research for more than 30 years have, however, represented managerial skills as two independent skill patterns, consideration and initiating structure, that have to be mastered concurrently for success (e.g. Blake and Mouton, 1964; Stogdill and Coons, 1957). These models also lack sufficient definitional detail to study managerial proficiency or the effectiveness of training (Morrison *et al.*, 1978; Schriesheim and Kerr, 1977). Yukl (1994) argues that further refinement of these constructs is needed by identifying specific skills making up each construct.

Some researchers through observational studies have attempted to identify the specific skills of managers (e.g. Cox and Cooper, 1988; Graham, 1983; Luthans *et al.*, 1985; Mitzberg, 1973; Stewart, 1967). These taxonomies fail to identify how the managerial skills interrelate (Yukl, 1994) and the conditions under which the skills need to be used. The competing values model (Quinn, 1988) and the managerial task cycle (Wilson *et al.*, 1990) are two theoretical

approaches that appear to overcome these criticisms. In addition, the second model has been compared to over 20 taxonomies of managerial skills and found to have broad theoretical support and high face validity as well as significantly to overlap in the identification of specific skills (Morrison *et al.*, 1978; Shipper, 1991). Thus, the managerial task cycle model was chosen to identify the specific managerial skills to study.

Method

Design

A pre- and post-test control group design was chosen for this field study because it is classified as an experimental design and allows for some degree of control of confounding variances (Campell and Stanley, 1963). In addition, the choice of this design is responsive to calls for improved evaluation of training (Brown, 1980; Dunn and Thomas, 1985).

Sample

The sample consisted of 60 middle managers employed by a private company contracted to operate four federally funded vocational training sites, called Job Corps Centers, for young adults 16 to 24 years of age. The four sites are widely dispersed geographically within the 48 contiguous states of the USA. Managers from two sites functioning at low and high performance levels, as determined by a measurement system established by the Department of Labor (DOL), were assigned to the control and experimental groups, respectively. Complete data were obtained on 26 managers in the control group and 27 in the experimental group.

The average individual was 44 years old and had 11 years of managerial experience. In addition, 42 per cent of the managers were female and 57 per cent male. Eighty per cent of the sample was white and 20 per cent non-white.

Measures

Managerial skills were assessed using a structured questionnaire, The Survey of Management Practices – Form K (Wilson and Wilson, 1991). It was chosen for two reasons: the first being its comprehensive nature and second, its psychometric soundness (Clark *et al.*, 1992; Koser and Lussier, 1987; Morrison *et al.*, 1978; Shipper, 1995; Van Velsor and Leslie, 1991). The survey of management practices (SMP) is the only instrument to have twice been judged by the Center for Creative Leadership to have acceptable psychometric properties (Morrison *et al.*, 1978; Van Velsor and Leslie, 1991). The questionnaire consists of 115 items of which 71 are purported to measure 13 managerial skills and are used in this study. The other items in the questionnaire measure attitudes.

Each skill scale consists of five to six items. Responses are recorded on a 7-point Likert scale ranging from an extremely high to an extremely low extent. Three typical items are as follows: this manager (supervisor, etc.) plans the work, so it keeps running smoothly, gets advice from the group on the best way

to do things, and compliments individuals who contribute significantly to the group's effort.

Morrison *et al.* (1978) rated an earlier version (Form G) to be one of the three best of 24 managerial behavior instruments reviewed. The instruments were reviewed on linkage to theory and research, and psychometric properties, among other criteria. Van Velsor and Leslie (1991) rated another version (Form J) of the SMP to be one of 16 acceptable managerial behavior instruments of 40 reviewed. The instruments were reviewed on psychometric properties and usefulness for performance enhancement through feedback.

Prior forms of the questionnaire have been examined for test/re-test reliability, internal consistency, interrater reliability, construct validity, and criterion validity (Shipper, 1995; Wilson, 1975, 1978; Wilson *et al.*, 1996). All of the scales within the instrument have been reported in prior studies to exceed Nunnally's (1978) criteria for reliability in exploratory research. All of the scales were retested for internal consistency in this study and again found to exceed Nunnally's criteria as reported in the next section. The scales were also tested for interrater consensus using the Interrater Agreement Index ($r_{WG(j)}$) (James *et al.*, 1984). All scales were found to have acceptable interrater agreement as discussed in the next section.

Another reason for choosing the SMP was that the instrument has theoretical and empirical support for being a measure of a manager's competency with managerial skills. In contrast, most of the instruments available to gather observations on managerial behaviors focus on frequency (Bernardin and Beatty, 1984; Schriesheim and Kerr, 1974; Shipper, 1991; Van Velsor and Leslie, 1991; Yukl, 1989, 1994). Competency and not frequency may be, however, the important attribute concerning subunit performance (Van Velsor and Leslie, 1991). The managerial skill scales of the SMP differ from many others in that each item was intended to focus on competency and not just frequency (Wilson and Wilson, 1991). The difference is subtle and complex (Van Velsor and Leslie, 1991). In the SMP the difference can be found in both the anchor and the stem for each item. For example, Van Velsor and Leslie (1991) suggest that the appropriate anchors for a competency-based scale are "to a great extent/to some extent/not at all" (p. 10). The anchors for the SMP, "extremely small extent, never, not at all" and "extremely great extent, always," parallel the suggested anchors. In contrast, the anchors for Campbell's (1991) Leadership Index (CLI) are "always" and "never," and the anchors for Bass's (1990) Multi-factor Leadership Questionnaire are "not at all" and "frequently, if not always."

In addition, Shipper (1991) reframed the SMP for frequency, correlated the reframed version with the original, and compared the two versions on criterion validity. He found that none of the shared variances of the paired scales exceeded 10 per cent and that, overall, competency was associated with managers of high-performing work units, whereas frequency was associated with managers of low-performing work units. Therefore, the SMP appears to primarily assess competency of managerial behaviors.

Since the individual of interest in this study is the manager, the average score for all employees who responded to the questionnaire within a work unit was

used as the measure for each scale, as has been done in prior studies (e.g. Atwater and Yammarino, 1992; Hagarty, 1974; Prussia and Kinicki, 1996). Theoretically, such aggregation is appropriate when studying managers because it reduces random error and perceptual differences among observations by others (Campion, 1988). Empirically, calculating the index of interrater agreement and eta squared (η^2) for each scale can evaluate the appropriateness of such aggregation (Prussia and Kinicki, 1996). The interrater agreement indexes and etas squared are reported in Table I. All the interrater agreement indexes were 0.86 or higher. All etas squared were significant at the 0.05 level with just two exceptions at time t_2 . Thus, the combined results suggest within-groups agreement with two minor exceptions and thus, aggregation is justified.

Scales	t_1 Interrater agreement			t_2 Interrater agreement		
	Cronbach's α	index $r_{WG(j)}$	Eta squared η^2	Cronbach's α	index $r_{WG(j)}$	Eta squared η^2
Clarification of goals and objectives	0.94 $n = 193$	0.95 $n = 53$	0.53 $n = 53$	0.96 $n = 180$	0.94 $n = 53$	0.37 $n = 53$
Upward communication	0.93 $n = 199$	0.93 $n = 53$	0.52 $n = 53$	0.94 $n = 186$	0.93 $n = 53$	0.48 $n = 53$
Orderly work planning	0.95 $n = 198$	0.93 $n = 53$	0.55 $n = 53$	0.95 $n = 189$	0.92 $n = 53$	0.40 $n = 53$
Organizational expertise	0.89 $n = 198$	0.93 $n = 53$	0.63 $n = 53$	0.93 $n = 190$	0.94 $n = 53$	0.47 $n = 53$
Functional expertise	0.89 $n = 197$	0.92 $n = 53$	0.60 $n = 53$	0.92 $n = 189$	0.93 $n = 53$	0.45 $n = 53$
Work facilitating	0.92 $n = 198$	0.90 $n = 53$	0.53 $n = 53$	0.92 $n = 188$	0.91 $n = 53$	0.39 $n = 53$
Methods management	0.92 $n = 197$	0.91 $n = 52$	0.55 $n = 53$	0.93 $n = 187$	0.94 $n = 53$	0.42 $n = 53$
Providing feedback	0.89 $n = 196$	0.87 $n = 53$	0.45 $n = 53$	0.91 $n = 188$	0.91 $n = 53$	0.37 $n = 53$
Time emphasis	0.89 $n = 199$	0.93 $n = 53$	0.52 $n = 53$	0.91 $n = 190$	0.93 $n = 53$	0.42 $n = 53$
Control of details	0.77 $n = 196$	0.90 $n = 53$	0.42 $n = 53$	0.85 $n = 187$	0.86 $n = 53$	0.32 n/s $n = 53$
Goal pressure	0.81 $n = 194$	0.90 $n = 52$	0.48 $n = 53$	0.82 $n = 186$	0.90 $n = 53$	0.33 n/s $n = 53$
Delegation (permissiveness)	0.90 $n = 199$	0.93 $n = 53$	0.46 $n = 53$	0.90 $n = 187$	0.92 $n = 53$	0.46 $n = 53$
Recognition for good performance	0.97 $n = 201$	0.91 $n = 53$	0.40 $n = 53$	0.97 $n = 184$	0.95 $n = 53$	0.39 $n = 53$

Note: All η^2 at t_1 significant at $p \leq 0.001$; all η^2 at t_2 significant at $p \leq 0.05$ except where noted

Table I.
Test of reliability of scales

Only the employees' observations of managerial skills were used in this study to avoid the obvious bias of self-reported observations (Tubbs, 1986). Furthermore, subordinate observations have previously been found more discriminating than either self or superior observations relative to effectiveness (Wilson *et al.*, 1990). Theoretically, subordinates normally have more opportunities to observe the array of managerial behaviors of interest than do superiors (Bernardin, 1986). Thus, the use of subordinate observations should ensure a more accurate evaluation of the effects of the training.

Procedures

Two statistical procedures were used to determine if the training was beneficial to the managers. The first procedure was to test each scale using a classic experimental analysis of variance (ANOVA). This procedure allowed for each skill measured to be tested for significant changes.

The second procedure was a series of four Wilcoxon matched-pairs signed-tests. The technique has been used before to test for differences in skill profiles (Shipper, 1991). This analysis tests the hypothesis that there is no difference between two profiles that consist of ordered-metric scales (SPSS, 1990). The test accounts for the magnitude and the direction of differences between the paired scales (Siegel, 1956). To determine if the training was effective four tests of statistical significance must be conducted. First, the profiles of the experimental and control groups at t_1 must be compared to check for initial comparability. Second, the profiles of the control groups at t_1 and t_2 must be compared to check for change due to extraneous variances not controlled for in the design of the experiment. Third, the profiles of the experimental groups at t_1 and t_2 must be compared to check for change due to either experimental or extraneous variance or both. Fourth, the profiles of the experimental and control groups at t_2 must be compared to test for change due to experimental variance. Only if the first two tests are nonsignificant and the second two are significant can one conclude that the effect is due to experimental variance. This procedure allows for the overall profiles to be tested for significant changes.

Intervention

All the managers in both the experimental and control groups received information on how to interpret the feedback that they had received from their peers, subordinates and superior. The control group received a one-day seminar specific to the results of the survey and the management theory that supported it but no management training. The experimental group received feedback and a training program that used individual/group processes and content formats. The context of the training was based on the Task Cycle Model of Managerial Behavior developed by Clark Wilson (1989). The program consisted of two four-hour sessions per month over a three-month period for a total of 24 hours of management training. Session I provided the study group participants with feedback and interpretation of the survey results and an overview of the Task Cycle Management Training Course similar to what the members of the control

group had received. Session II focused on the definition and purpose of goals and how managers can benefit from clear goal development. The levels of planning and how planning, problem solving, and decision making relate to each another were skill areas covered in Session III. The next session introduced the coaching process, developing and using coaching guides, and how coaching strengthens performance reviews. Session V involved the appraisal/counseling process and its effects on employee performance. The final session dealt with exercising positive management control through quality circles and participative management techniques and motivating others to work for them through reinforcement and expectancy of success strategies[1].

Results

The results of the comparison of each skill of the managers in control and experimental groups at time t_1 were all nonsignificant as one would expect (Table II). These results suggest that the groups were equivalent at time t_1 although the managers were not randomly assigned to the two groups.

The results of the analysis of variance on each scale were not supportive that significant changes had taken place for any managerial scales for the main effect due to training at the normal level of statistical significance of 0.05 (Table III). An indicated, training main effect was found at the statistical significant level of 0.10 for one skill, goal pressure. Similarly, the results of the analysis of variance on each scale were not supportive that significant changes had taken place for any managerial scales for a pre-post main effect at the normal level of statistical significance of 0.05. In contrast, an indicated pre-post main effect was found at the statistical significant level of 0.10 for five of the 13 managerial skills.

The results of the analysis of profiles, however, yielded results at normal levels of statistical significance (Table IV). The first comparison was another

Scales	Control group $n = 26$ \bar{x}	Experimental group $n = 27$ \bar{x}	Sig. p
Clarification of goals and objectives	4.95	5.16	0.471
Upward communication	5.02	5.00	0.928
Orderly work planning	4.84	4.83	0.983
Organizational expertise	5.30	5.41	0.719
Functional expertise	5.11	5.10	0.988
Work facilitating	4.91	4.98	0.798
Methods management	4.94	5.13	0.507
Providing feedback	4.77	4.92	0.569
Time emphasis	5.37	5.53	0.490
Control of details	4.13	4.28	0.448
Goal pressure	3.36	3.09	0.285
Delegation (permissiveness)	5.25	5.23	0.933
Recognition for good performance	5.16	5.25	0.741

Table II.
Test for equivalency
between control and
experimental groups at t_1

Scales	Control group t_1 $n = 26$	Experimental group t_1 $n = 27$	Control group t_2 $n = 26$	Experimental group t_2 $n = 27$	Pre-post significance ρ	Training significance ρ
Clarification of goals and objectives	4.95	5.16	5.26	5.54	0.055	0.182
Upward communication	5.02	5.00	4.96	5.35	0.435	0.353
Orderly work planning	4.84	4.83	5.06	5.30	0.086	0.556
Organizational expertise	5.31	5.41	5.39	5.72	0.291	0.241
Functional expertise	5.11	5.10	5.21	5.52	0.168	0.428
Work facilitating	4.91	4.98	5.18	5.44	0.055	0.370
Methods management	4.94	5.13	5.20	5.48	0.092	0.198
Providing feedback	4.77	4.92	4.98	5.34	0.060	0.131
Time emphasis	5.37	5.53	5.37	5.66	0.666	0.115
Control of details	4.13	4.28	4.26	4.40	0.381	0.296
Goal pressure	3.36	3.09	3.41	3.14	0.750	0.087
Delegation (permissiveness)	5.25	5.23	5.09	5.37	0.986	4.10
Recognition for good performance	5.16	5.25	5.14	5.54	0.470	0.203

Table III.
Results of ANOVA for
effects due to training

Notes:

The highest cell mean for each scale appears in bold

None of the interactions effects approached significance therefore they were omitted

check to determine if the control and experimental groups were equivalent at time t_1 . The results were nonsignificant suggesting that they were. The second comparison was for the profiles of the control group at times t_1 and t_2 to ascertain if any intervening variable that had not been controlled had caused a change in the profiles. The results were significant suggesting that some uncontrolled variable had intervened.

The third comparison was of the profiles of the experimental group at times t_1 and t_2 to ascertain if a significant change had taken place. The results suggest that a change had taken place. The fourth comparison was between the control and experimental group at time t_2 to ascertain if the latter's profile was significantly different from the former's and in the expected direction. The results were supportive that the change in the experimental group's profile was significantly different than the control group's and in the appropriate direction.

Discussion

The purpose of this research was to test the effectiveness of training in a management development program using 360 feedback. The results of the study indicate partially that training enhanced the effectiveness of the 360 feedback in the development of managerial skills. This indication is at the overall profile level and not at the individual skill level. A possible reason for this finding is that the emphasis of the Clark Wilson Task Cycle Model on which the training was based is the development of a balanced profile (Wilson and

Comparison between profiles of control and experimental groups	Mean rank	Cases
t_1	4.60	5 – ranks (Et_1 LT Ct_1)
	8.50	8 + ranks (Et_1 GT Ct_1)
		0 ties (Et_1 EQ Ct_1)
		13 Total
	$Z = -1.5724$	2-tailed $P = 0.1159$
t_1 and t_2	3.67	3 – ranks (Ct_2 LT Ct_1)
	7.44	9 + ranks (Ct_2 GT Ct_1)
		1 ties (Ct_2 EQ Ct_1)
		13 Total
	$Z = -2.1965$	2-tailed $P = 0.0281$
t_1 and t_2	0	0 – ranks (Et_2 LT Et_1)
	7.00	13 + ranks (Et_2 GT Et_1)
		0 ties (Ct_2 EQ Ct_1)
		13 Total
	$Z = -3.1798$	2-tailed $P = 0.0015$
t_2	4.00	1 – ranks (Et_2 LT Ct_2)
	7.25	12 + ranks (Et_2 GT Ct_2)
		0 ties (Et_2 EQ Ct_2)
		13 Total
	$Z = -2.9003$	2-tailed $P = 0.0037$

Table IV.
Analysis of pre (t_1) and
post (t_2), control (C) and
experimental (E) profiles
using Wilcoxon
matched-pairs signed-
rank test

Shipper, 1992). Thus, a participant in the training would be urged to first develop the skill of which they have least. Assuming that the participants had different profiles initially, different participants would be urged to develop different skills than other participants. Therefore, the increase in any one skill for all the participants would not be as recognizable as the increase in the profile of all the participants.

A second possible explanation as to why the training effect was not found to be more significant can be detected by examining the differences in eta squared between t_1 and t_2 . On each of the 13 scales eta squared dropped between t_1 and t_2 . Eta squared is a measure of within-group consensus (cf. Riggs and Knight, 1994). Thus, the employees of the managers had stronger consensus regarding the manager's skills at t_1 than at t_2 . This finding may be an artifact of some employees' subconscious unwillingness to acknowledge the change in the manager's skill level through selective perception. In other words, some employees may be selectively perceiving only the manager's behaviors that reinforce their observations at t_1 during the experimental time period. This finding would suggest that for some employees the old adage that first impressions are hard to change could also be true for them relative to their

perceptions of managerial skills during a one-year time period. Thus, it would appear that managers must be persistent and consistent in exhibiting their new skills if they want their employees to be convinced that their skill level has really changed.

Another possible explanation as to why the training effect was not found more significant than it was could be due to individual initiative of the participants in the control groups. Although the managers of the participants in the control group agreed to not provide any training during the experimental period, nothing was done to stop the individuals from seeking to improve themselves once they received the feedback. Follow-up interviews indicated that at one site where approximately half the control group was located some participants had taken it on themselves to improve. The statistical analysis of the pre and post profiles of the control group supports that the skills of the managers of that group did improve. Why this happened cannot be said with certainty, but any improvement in the control group would lessen the chances of finding statistically significant changes in the experimental group due to training. At the same time, the comparisons made of the group profiles indicate that significantly larger changes occurred in the experimental group than in the control group during the study suggesting that the training was effective at enhancing development of the managerial skills.

Limitations of the study

An obvious limitation of this study is that the experimental design relative to 360 feedback is a one-group pretest-post-test design. This design is called pre-experimental (Campbell and Stanley, 1963). This limitation was accepted for two reasons. First, the target variable of interest was training. The research design for measuring the impact of the training included both an experimental and control group with pre and post measures. Thus, for measuring the impact of training a true experimental design was used. Second, an ethical and pragmatic issue arose. To conduct a true test of 360 feedback would have required two additional groups and the withholding of information collected from the survey for over a year. Participants want to know the results (Dunham and Smith, 1979). Some would argue that the act of withholding the results of the survey is unethical because it violates the needs of the participants and could contaminate attitudes within the organization against future surveys. In addition, failure to provide feedback could create turmoil within the organization that the managers were not willing to risk. Thus, design of this study with its limitation was chosen because the primary variable of interest (training) was incorporated in an experimental design although another variable of interest (receiving 360 feedback) was not.

Another limitation of this study is the sample size. The actual control and experimental groups once the data were collected were only 26 and 27, respectively. Other studies using subordinate feedback have also been plagued by this limitation (Church, 1997). This limited the ability to find statistical significance at the individual skill level. One possible reason for small sample

sizes could be both the direct and indirect costs of conducting such studies. Church (1997), however, indicated that more research is needed with diverse and larger samples to both understand the multirater process and justify the use of it. This study at least provides some support for both additional research and for the need for training when using 360 feedback to obtain optimal results.

A third limitation of this study was the inability to measure actual results of changes in performance in the managers as part of the pre-post assessment. Kirkpatrick (1976) identifies this as the highest of four levels of rigor in assessing the effectiveness of training. This study focussed on the third level of rigor, the participants' behavior. A prior study based on the same model simultaneously measured both levels of rigor and found significant positive changes in both (Shipper and Neck, 1990). Thus, some evidence exists to support the expectation that if performance had been measured, improvements would have been found. However, this is speculative and future research should include the fourth level of rigor.

A final and obvious limitation to this study is that it investigates only one training program that could be developed from 360 feedback. Other training programs could be developed that could be more or less effective. Additional research is needed to learn the most effective training to be used in conjunction with 360 feedback.

Note

1. A more detailed description of the training is available from the authors.

References

- Atwater, L.E. and Yammarino, F.J. (1992), "Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions?", *Personnel Psychology*, Vol. 45, pp. 141-64.
- Bass, B.M. (1990), *Bass & Stogdill's Handbook of Leadership: Theory, Research and Managerial Applications*, 3rd ed., Free Press, New York, NY.
- Bennis, W.G., Benne, K.D. and Chin, R. (1969), *The Planning of Change*, 2nd ed., Holt, Rinehart and Winston, New York, NY.
- Bernardin, H.J. (1986), "Subordinate appraisal: a valuable source of information about managers", *Human Resource Management*, Vol. 25, pp. 421-39.
- Bernardin, H.J. and Beatty, R.W. (1984), *Performance Appraisal: Assessing Human Behavior at Work*, Kent Publishing, Boston, MA.
- Blake, R.R. and Mouton, J.S. (1964), *The Managerial Grid*, Gulf, Houston, TX.
- Brown, M.G. (1980), "Evaluating training via multiple baseline designs", *Training and Development Journal*, Vol. 34 No. 10, October, pp. 11-16.
- Campbell, D. (1991), *Manual for the Campbell Leadership Index*, National Computer Systems, Minneapolis.
- Campbell, D.T. and Stanley, J.C. (1963), *Experimental and Quasi-Experimental Designs for Research*, Rand McNally, Chicago, IL.
- Campion, M.A. (1988), "Interdisciplinary approaches to job design: a constructive replication with extensions", *Journal of Applied Psychology*, Vol. 78 No. 3, pp. 467-81.
- Carnevale, A.P., Schultz (1990), "Return on investment: accounting for training", *Training & Development Journal*, Vol. 44 No. 7, July, pp. 41-72.

- Church, A.H. (1997), "Managerial self-awareness in high-performing individuals in organizations", *Journal of Applied Psychology*, Vol. 82 No. 2, pp. 281-92.
- Clark, K.E., Clark, M.B. and Campbell, D.P. (1992), *Impact of Leadership*, Center for Creative Leadership, Greensboro, NC.
- Cox, C.J. and Cooper, C.L. (1988), *High Flyers: An Anatomy of Managerial Success*, Blackwell, New York, NY.
- Dunham, R.B. and Smith, F.J. (1979), *Organizational Surveys: An Internal Assessment of Organizational Health*, Scott, Foresman and Company, Glenview, IL.
- Dunn, S. and Thomas, K. (1985), "Evaluating training via multiple baseline designs", *Training*, Vol. 22, October, pp. 28-40.
- Fayol, H. (1949), *General and Industrial Management*, C. Storrs (trans), Pitman, London.
- Graham, J.L. (1983), "Brazilian, Japanese, and American business negotiations", *Journal of International Business Studies*, January, Vol. 16, pp. 81-96.
- Hagerty, W.H. (1974), "Using subordinate ratings to elicit behavioral changes in supervisors", *Journal of Applied Psychology*, Vol. 59, pp. 764-6.
- James, L.R., Demaree, R.G. and Wolf, G. (1984), "Estimating with in-group interrater reliability with and without response bias", *Journal of Applied Psychology*, Vol. 69, pp. 85-98.
- Kirkpatrick, D. (1976), "Evaluation of training", in Craig, R.L. (Ed.), *Training and Development Handbook*, McGraw-Hill, New York, NY.
- Koser, K. and Lussier, R.R. (1987), *A Summary of the Research Related to Middle Management*, Veterans Administration Medical Center, Long Beach.
- Lewin, K. (1948), *Resolving Social Conflicts*, Harper & Bros, New York, NY.
- Luthans, F., Rosenkrantz, S.A. and Hennessey, H.W. (1985), "What do successful managers really do? An observational study of managerial behaviors", *Journal of Applied Behavioral Science*, Vol. 21, pp. 255-70.
- Mintzberg, H. (1973), *The Nature of Managerial Work*, Harper & Row, New York, NY.
- Morrison, A.M. and McCall, J. DeV. (1978), *Feedback to Managers: A Comprehensive Review of Twenty-Four Instruments*, Center for Creative Leadership, Greensboro, NC.
- Nunnally, J.C. (1978), *Psychometric Theory*, McGraw-Hill, New York, NY.
- Porter, L.W. and McKibbin, L.E. (1988), *Management Education and Development: Drift or Thrust into the 21st Century?*, McGraw-Hill, New York, NY.
- Prussia, G.E. and Kinicki, A.J. (1996), "A motivational investigation of group effectiveness using social-cognitive theory", *Journal of Applied Psychology*, Vol. 81, pp. 187-98.
- Quinn, R.E. (1988), *Beyond Rational Management*, Jossey-Bass, San Francisco, CA.
- Riggs, M.L. and Knight, P.A. (1994), "The impact of perceived group success-failure on motivational beliefs and attitudes: a causal model", *Journal of Applied Psychology*, Vol. 79, pp. 755-66.
- Schriesheim, C.A. and Kerr, S. (1974), "Psychometric properties of the Ohio State leadership scales", *Psychological Bulletin*, Vol. 81, pp. 756-65.
- Schriesheim, C.A. and Kerr, S. (1977), "Theories and measurement of leadership: a critical appraisal of present and future directions", in Hunt, J.G. and Larson, L.L. (Eds), *Leadership: The Cutting Edge*, Southern Illinois University Press, Carbondale, IL.
- Shipper, F. (1991), "Mastery and frequency of managerial behaviors relative to sub-unit effectiveness", *Human Relations*, Vol. 44 No. 4, pp. 371-88.
- Shipper, F. (1995), "A study of the psychometric properties of the managerial skill scales of the survey of management practices", *Educational and Psychological Measurement*, Vol. 55 No. 3, pp. 468-79.
- Shipper, F. and John, J. (1992), "Employees' feedback: its use for management development and the results in a government organization", in Fargher, J.S. (Ed.), *Proceedings of Symposium on*

- Productivity and Quality Improvement with a Focus on Government*, Industrial Engineering and Management Press, Washington, DC.
- Shipper, F. and Neck, C.P. (1990), "Subordinates' observations: feedback for management development", *Human Resource Development Quarterly*, Vol. 1 No. 4, pp. 371-84.
- Shore, T.H., Shore, L.M. and Thornton, I.G.C. (1992), "Construct validity of self- and peer evaluations of performance dimensions in an assessment center", *Journal of Applied Psychology*, Vol. 77, pp. 42-54.
- Siegel, S. (1956), *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, New York, NY.
- SPSS Reference Guide* (1990), SPSS, Chicago.
- Stephen, E., Mills, G.E., Pace, R.W. and Ralphs, L. (1988), "HRD in the Fortune 500", *Training and Development Journal*, January, pp. 26-32.
- Stewart, R. (1967), *Managers and their Jobs: A Study of the Similarities and Differences in the Way Managers Spend their Time*, Macmillan, London.
- Stogdill, R.M. and Coons, A.E. (1957), *Leader behavior: Its Description and Measurement*, Bureau of Business Research, Ohio State University, Columbus.
- Tubbs, M.E. (1986), "Goal setting: a meta-analytical examination of the empirical evidence", *Journal of Applied Psychology*, Vol. 71, pp. 474-83.
- Urwick, L.F. (1952), *Notes on the Theory of Organizations*, American Management Association, New York, NY.
- Van Velsor, E. and Leslie, J.B. (1991), *Feedback to Managers, Volume II: A Review and Comparison of Sixteen Multi-rater Feedback Instruments*, Center for Creative Leadership, Greensboro, NC.
- Whitley, R. (1989), "On the nature of managerial tasks and skills: their distinguishing characteristics and organization", *Journal of Management Studies*, Vol. 26, March, pp. 209-24.
- Wilson, C.L. (1975), *Multi-level Management Surveys: Feasibility Studies and Initial Applications*, *JSAS: Catalog of Selected Documents in Psychology*, Vol. 5, Ms. No. 1137, American Psychology Association, Washington, DC.
- Wilson, C.L. (1978), *The Wilson Multi-level Management Surveys: Refinement and Replication of the Scales*, *JSAS: Catalog of Selected Documents in Psychology*, Vol. 8, Ms. No. 1707, American Psychology Association, Washington, DC.
- Wilson, C.L. (1989), "Task cycle theory: a learning-based view of organization behavior", in Ochsman, R. (Ed.), *Psychology and Productivity*, Plenum Books, New York, NY.
- Wilson, C.L. and Shipper, F. (1992), *Task Cycle Management: A Competency-Based Course for Operating Managers*, rev. ed., The Clark Wilson Group, Silver Spring, MD.
- Wilson, C.L. and Wilson, J.L. (1991), *Teams and Leaders: A Manual for the Clark Wilson Publishing Company Training and Development Programs*, Clark Wilson Group, Silver Spring, MD.
- Wilson, C.L., O'Hare, D. and Shipper, F. (1990), "Task cycle theory: the process of influence", in Clark, K.E. and Clark, M.B. (Eds), *Measures of Leadership*, Leadership Library of America, West Orange, NJ, pp. 185-204.
- Wilson, C.L., Wilson, J.L. and Wilson, K.B. (1996), *Meaningful Measures*, Clark Wilson Group, Silver Spring, MD.
- Yukl, G.A. (1989), *Leadership in Organizations*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.
- Yukl, G.A. (1994), *Leadership in Organizations*, 3rd ed., Prentice-Hall, Englewood Cliffs, NJ.
- Zemke, R. and Zemke, S. (1995), "Adult learning: what do we know for sure?", *Training*, June, Vol. 32, pp. 31-7.

